The EMBO Journal Vol.16 No.4 pp.856–865, 1997

Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors

Deanna Kulpa, Robert Topping and Alice Telesnitsky[1]

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620, USA

[1]Corresponding author

Reverse transcriptase must perform two specialized template switches during retroviral DNA synthesis. Here, we used Moloney murine leukemia virus-based vectors to examine the site of one of these switches during intracellular reverse transcription. Consistent with original models for reverse transcription, but in contrast to previous experimental data, we observed that this first strand transfer nearly always occurred precisely at the 5′ end of genomic RNA. This finding allowed us to use first strand transfer to study the classes of errors that reverse transcriptase can and/or does make when it switches templates at a defined position during viral DNA synthesis. We found that errors occurred at the site of first strand transfer ~1000-fold more frequently than reported average reverse transcriptase error rates for template-internal positions. We then analyzed replication products of specialized vectors that were designed to test possible origins for the switch-associated errors. Our results suggest that at least some errors arose via non-templated nucleotide addition followed by mismatch extension at the point of strand transfer. We discuss the significance of our findings as they relate to the possible contribution that template switch-associated errors may make to retroviral mutation rates.

Keywords: fidelity/retrovirus/reverse transcriptase

Introduction

Models for reverse transcription propose that reverse transcriptase must perform two specialized template switches, known as 'strand transfers' or 'jumps', in order to complete the synthesis of the characteristic two-LTR (long terminal repeat) form of retroviral DNA (Figure 1) (Coffin, 1979; Gilboa *et al.*, 1979). During reverse transcription, the synthesis of one of two discrete-length DNA intermediates has been proposed to precede each strand transfer. This study focuses on the first strand transfer and minus strand strong stop DNA (–ssDNA), the putative intermediate that precedes it.

In the first strand transfer, reverse transcriptase is believed to switch from a donor template region (termed 'r') at the 5′ end of genomic RNA to an identical r region repeat at the genome's 3′ end (Coffin *et al.*, 1978; Swanstrom *et al.*, 1981). Viral RNA repeats are necessary for the first strand transfer, but it is conceivable that the first strand transfer may occur after only a portion of the r repeat is reverse transcribed. The r regions of various retroviruses differ substantially in length, suggesting that even the shortest natural r [which, at 12 nucleotides, is much shorter than the 68 nucleotide long r of Moloney murine leukemia virus (M-MuLV)] provides sufficient template to promote the first strand transfer (Coffin, 1996).

The prevailing model for reverse transcription suggests that the first strand transfer occurs only after –ssDNA synthesis is completed when reverse transcriptase reaches the 5′ end of genomic RNA (Gilboa *et al.*, 1979). This model was based in part on the observation that full-length –ssDNA is a prominent product in so-called 'endogenous reactions', which involve studying DNA synthesized after the addition of nucleotide substrates to detergent-permeabilized purified virions (Haseltine *et al.*, 1979). However, –ssDNA is not detectable in infected cells unless the infecting virus is defective in RNase H activity (Coffin, 1979; Blain and Goff, 1995). The apparent absence of –ssDNA from infected cells suggests either that full-length –ssDNA is short lived and all –ssDNA that is synthesized performs the first jump, or else that discrete-length –ssDNA is not formed during intracellular reverse transcription, possibly because the first strand transfer occurs before reverse transcriptase reaches the 5′ end of the RNA. Whether the first strand transfer occurs from the 5′ end of the genome or instead takes place from an earlier, internal position has different implications regarding which contacts between enzyme and primer–template are important during template switching. Footprinting of reverse transcriptase on a simple template has revealed that the enzyme makes extensive contacts with the template strand both in front of and behind the growing point for DNA synthesis, but what contacts are made in the template switch intermediate is unknown (Woehrl *et al.*, 1995a,b).

Two previous studies have examined the site of first strand transfer during viral replication. These studies tested a prediction of reverse transcription models: namely, that if one of the two r repeats of a viral RNA were altered genetically so that it differed from the other, then the sequence of the 3′ r (the acceptor template region) should be lost and that of the 5′ r (donor template) should be found in both DNA strands of both product LTRs. This predicted pattern of r inheritance is presented in Figure 1. Both previous studies that addressed the site of first strand transfer made use of viral templates with r region linker insertion mutations and both reports obtained the unpredicted finding that 3′ r sequences were frequently inherited. These results suggest that the first strand transfer may frequently if not always occur prematurely, before the 5′ end of RNA is reached (Lobel and Goff, 1985; Ramsey and Panganiban, 1993; Temin, 1993). A third report provided evidence that the first strand transfer often

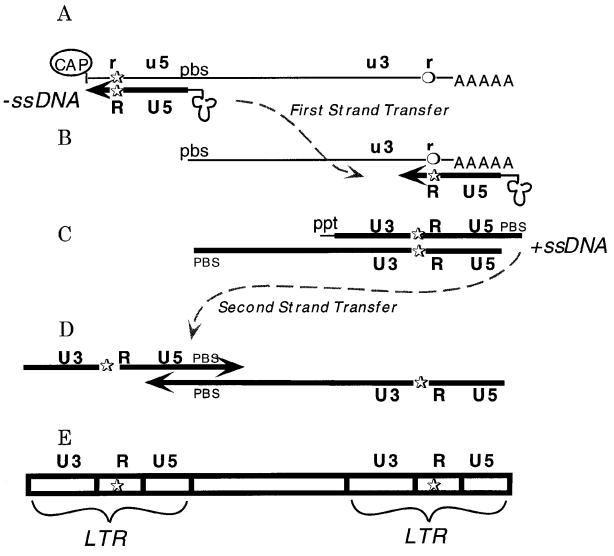

Fig. 1. Model for retroviral reverse transcription. (**A**) Minus strand DNA synthesis initiates from a tRNA primer partially annealed to the pbs (primer binding site) region of the plus sense RNA genome and proceeds to the 5′ end of genomic RNA, thereby generating minus strand strong stop DNA (–ssDNA). (**B**) Following RNase H degradation of r and u5 regions of the resulting RNA–DNA duplex, –ssDNA undergoes the first strand transfer to the 3′ end of genomic RNA, where –ssDNA R sequences anneal to complementary 3′ r sequences. Minus strand DNA synthesis resumes, accompanied by RNase H digestion of the template strand. (**C**) A short oligoribonucleotide that is called the ppt (polypurine tract) persists and primes plus strand DNA synthesis. Plus strand synthesis is halted after the portion of the tRNA primer that was originally annealed to the pbs is copied, thereby generating a DNA called plus strand strong stop DNA (+ssDNA). RNase H removes the RNA primers, and complementary PBS sequences at each end of the replication intermediate are exposed. (**D**) The annealing of these repeats constitutes the second strand transfer. Plus and minus strand syntheses are then completed (**E**), with each of the two DNA strands serving as template for the other's completion. The completed double-stranded DNA product contains two identical long terminal repeats (LTRs), each of which consists of the sequence elements U3, R and U5. In this figure, the star and the circle symbols denote sequence differences between genomic 5′ and 3′ r regions. Light lines and lower case letters = RNA; bold lines and upper case letters = DNA; CAP = 7-methyl-G cap nucleotide.

occurs within 23 nucleotides of the 5′ end of human immunodeficiency virus (HIV)-1 RNA (Klaver and Berkhout, 1994), but no systematic studies of the spectrum of positions from which first strand transfer occurs nor of the frequency of usage of these positions have been reported.

There has been some doubt as to whether full-length –ssDNA is competent to undergo the first strand transfer. The –ssDNA formed in endogenous reactions may be one nucleotide longer than its template, implying that a non-templated nucleotide may have been added (Swanstrom *et al.*, 1981). If this extended –ssDNA were to jump, then the 3′ end of –ssDNA would be non-complementary to the acceptor template region (the U3–R boundary) unless the added nucleotide happened to be complementary to the acceptor template. However, studies with purified enzymes show that reverse transcriptase is fairly efficient at extending primer-terminal mismatches (Perrino *et al.*, 1989). An important component of reverse transcriptase's high error rate is believed to be its efficiency at extending mismatches, a phenomenon observed both in reconstituted reactions and during viral replication (Perrino *et al.*, 1989; Roberts *et al.*, 1989; Peliska and Benkovic, 1992; Pulsinelli and Temin, 1994; Zinnen *et al.*, 1994; Das and Berkhout, 1995; Preston and Dougherty, 1996).

In this report, we generated a series of M-MuLV-based retroviral vectors with point mutations in their 3′ r regions which permitted us to examine the site of first strand transfer during intracellular replication. We also studied the errors which reverse transcriptase commits when it performs the first strand transfer *in vivo*.

Results

Vectors to examine the position of first strand transfer

We generated a series of replication-defective retroviral vectors that allowed us to determine the position of first strand transfer from an examination of product DNAs. Our vectors were encoded by derivatives of pBabe puro, a retroviral vector plasmid that contains the puromycin resistance gene (Morgenstern and Land, 1990). In each of our vectors, the two r regions differed from each other by a single base change: the 5′ (donor) r remained wild-type while a point mutation was introduced into the 3′ (acceptor) r. Each vector's mutation either destroyed an existing restriction enzyme recognition sequence or else introduced a new site (Figure 2). The rationale for using these mutant vectors was as follows: if first strand transfer occurred from a position in the donor r that was 'before' the site of a mutation in the acceptor r, then the progeny DNA would acquire the sequence of the acceptor r. If strand transfer occurred from a position 'after' the site of a mutation, then viral DNA would possess the sequence of the donor r (Figure 2). Hence, the restriction pattern of a product DNA would indicate where first strand transfer had occurred during its synthesis.

Examination of the site of strand transfer

To generate reverse transcription products, vector plasmids were stably transfected into ecotropic packaging cells, Rat2 cells were infected with vector-containing virions harvested from the vector-producing cells, and low molecular weight DNA (which included unintegrated viral DNA) was harvested from the infected cells. We used a PCR-based assay to analyze progeny DNAs because this allowed us to examine large populations of reverse transcription products and avoid sampling biases. We estimated what portion of progeny DNAs arose from strand transfer at various positions by digesting the PCR products with appropriate restriction enzymes (Figure 3), and we verified our findings by examining the prevalence of classes of progeny DNAs in individual clones. We examined the inheritance of sequences in r that were 1, 5, 9 and 14 nucleotides from the U3–R boundary (designated positions –1, –5, –9 and –14).

Analysis of acceptor r-mutant vector products demonstrated that the first strand transfer occurred almost exclusively at the U3–R boundary. The mutations of the –1 and –5 vectors destroyed an *Asc*I restriction site that straddles the U3–R boundary. If strand transfer occurred after the completion of –ssDNA, these vectors' progeny DNAs would possess *Asc*I sites; they would not be cleavable by *Asc*I if strand transfer occurred prematurely.

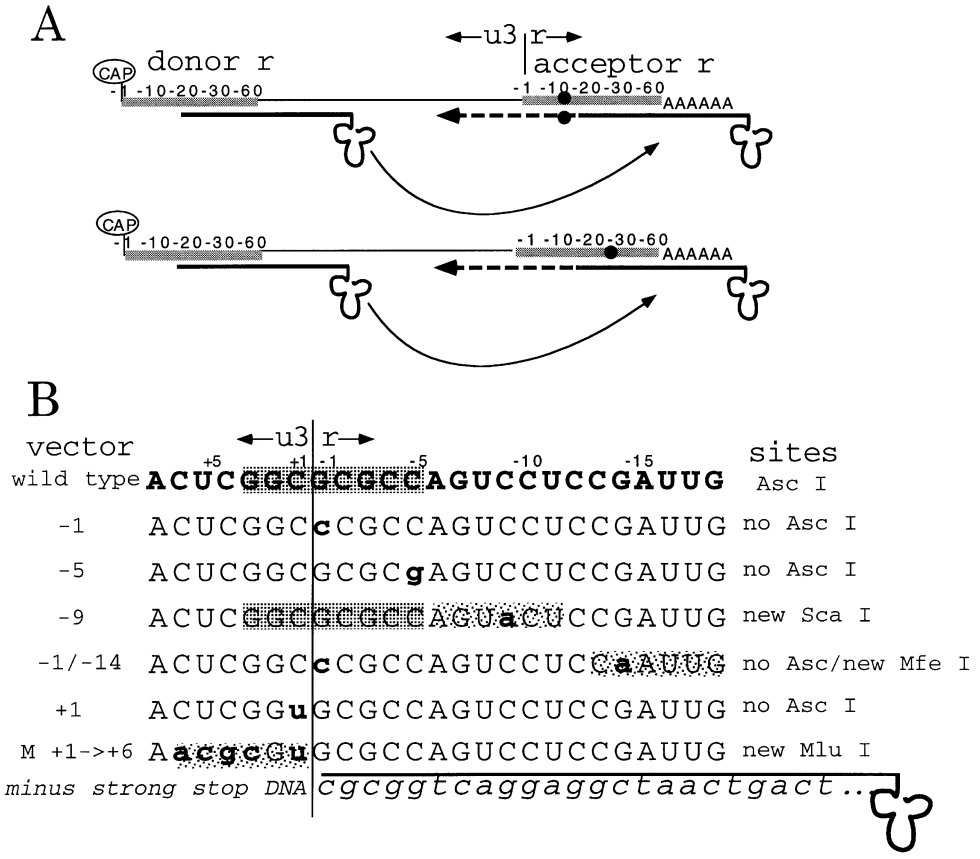

Fig. 2. Assay for site of first strand transfer. (**A**) Assay rationale: if a minus strand DNA product (solid line) strand-transfers to a mutagenized 3′ acceptor r before copying the donor r past the site of the mutation (indicated by heavy dot) in the 3′ r, then the mutation will be copied into product DNA, as indicated by the presence of the mutation (heavy dot) within the sequence of the DNA whose synthesis was templated after strand transfer (dotted line). If the first strand transfer occurs after the mutation site, product DNAs will bear the sequence of the 5′ donor r, as indicated by the absence of the mutation (heavy dot) in the DNA strand. (**B**) Vectors used to assess the position of transfer. Sequences of acceptor template u3–r junctions. The *Asc*I site in the wild-type vector is shaded and the newly introduced restriction sites in the other vectors are stippled. Nucleotides in the test vectors that differ from the wild-type sequence are presented in bold lower case letters. The sequence of completed –ssDNA is indicated at the bottom of the figure for reference.

We observed no decrease in the amount of *Asc*I digestion for the progeny of the –5 and –1 mutants as compared with DNA products of a vector containing wild-type sequences in both its r regions (henceforth 'the wild-type vector') (Figure 3B). Therefore, it appeared that for most DNA products, –ssDNA synthesis was completed prior to the first strand transfer. This was true even in the case of the –1 mutant, for which strand transfer at the U3–R junction required elongation of a mismatched primer terminus. Observations with –9 and –14 vectors were consistent with the conclusion that strand transfer generally occurred at the U3–R junction. Only very low levels of strand transfer prior to –14 were detectable by restriction analysis of progeny DNAs templated by a vector with mutations at both –1 and –14, and the site that would have been gained by premature strand transfer on a vector with an acceptor r mutation at –9 was not detected by restriction analysis.

Identification of strand transfer-associated errors

An unexpected finding of these studies was that a significant portion of progeny DNAs from the wild-type vector had lost the wild-type U3–R boundary sequence. Models for reverse transcription predict that, except for rare point mutations which might occur at a single position in roughly one out of 10^4 progeny DNAs, the U3–R sequence should not change through a single cycle or even multiple cycles of reverse transcription, regardless of the site of first strand transfer.

Alteration of the U3–R boundary sequence became evident when PCR products of wild-type vector progeny were found to display an unanticipated restriction digestion pattern. These DNAs should be fully digestible by *Asc*I, the restriction enzyme whose site straddles the U3–R junction, regardless of the site of strand transfer. However, a significant proportion (~10%) of these DNAs reproducibly failed to be cleaved by *Asc*I (Figure 3). Although some of the uncleaved product may have resulted from incomplete digestion, when undigested products were excised from gels and re-amplified, the viral amplified undigested products were largely resistant to *Asc*I cleavage whereas amplified undigested products from proviral plasmids were digested essentially to completion by *Asc*I (data not shown). Hence, the loss of the *Asc*I site among ~10% of the amplified wild-type viral DNA products did not appear to be due to errors by *Taq* DNA polymerase or other aspects of the PCR process, since repeated rounds of amplification of pNCA provirus plasmid DNA showed

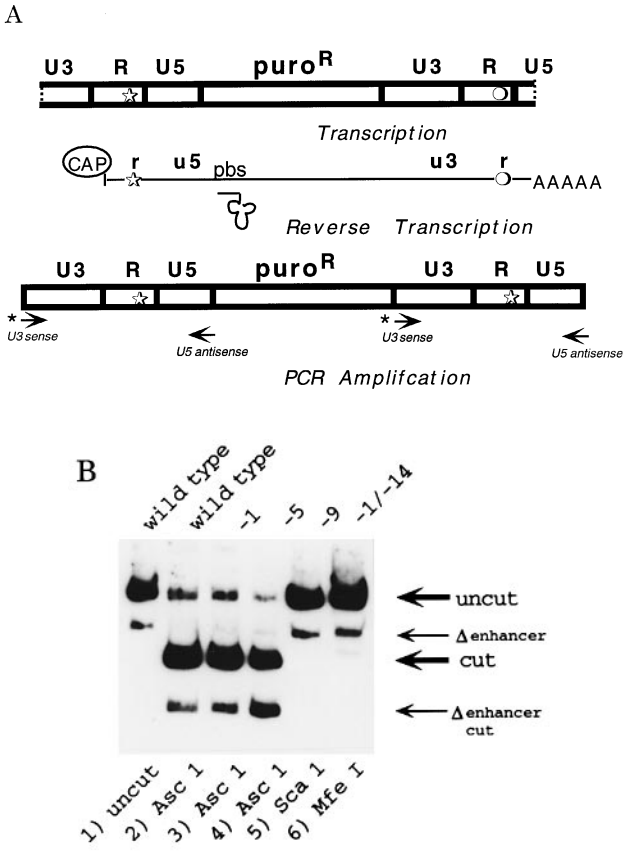

Fig. 3. Analysis of progeny DNA PCR products. (**A**) Experimental scheme. (i) Tipless provirus plasmids were engineered so that their two R regions differed by single point mutations (indicated with a star and a circle). The point differences were designed to either introduce a new or else destroy an existing restriction enzyme recognition site. (ii) Transcription of these proviruses within transfected 3T3 cells yielded viral RNAs that also had two different r regions. (iii) Reverse transcription of these RNAs generated DNAs with two intact viral LTRs. LTR sequences were PCR amplified using U3 sense and U5 antisense primers, one of which was radiolabeled (indicated by *). (**B**) Restriction analysis of progeny DNA PCR products. Radiolabeled LTR-containing PCR products of DNAs generated by vectors shown in Figure 2 were digested with the restriction enzymes indicated at the bottom of the figure. Arrows indicate the mobilities of the intact PCR products and of the radiolabeled fragment of restriction enzyme-digested ('cut') PCR products. Δ enhancer indicates the mobility of products that had lost one copy of the M-MuLV U3 region 72 bp repeat, presumably via homologous recombination during reverse transcription (Hu *et al.*, 1993). Lane 1, wild-type vector undigested; lane 2, wild-type vector PCR products digested with *Asc*I; lane 3, –1 vector products digested with *Asc*I; lane 4, –5 products digested with *Asc*I; lane 5, –9 products digested with *Sca*I; lane 6, –1/–14 vector products digested with *Mfe*I.

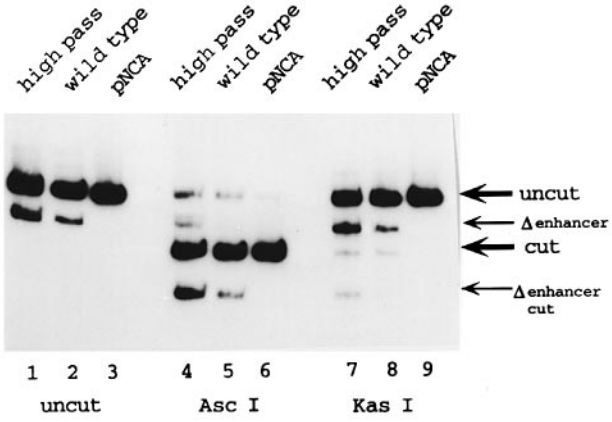

Fig. 4. Restriction digestion assay examining the frequency of +1G mutation. The origin of PCR-amplified products is indicated at the top of the figure. Wild-type are amplified reverse transcription products of the wild-type vector; 'high pass' indicates reverse transcription products of high serial passage infectious M-MuLV cultured as described in Materials and methods; pNCA indicates amplified M-MuLV provirus DNA templated by pNCA (Colicelli and Goff, 1988), a provirus plasmid with intact LTRs. Lanes 1–3, uncleaved PCR products; lanes 4–6, PCR products cleaved with *Asc*I; lanes 7–9, PCR products cleaved with *Kas*I. Mobility designations are as in Figure 3.

no evidence of *Asc*I site loss. In the experiments presented here, we did not examine the significance of vector-to-vector differences in apparent rates of *Asc*I site loss.

To examine what alterations prevented *Asc*I cleavage of some viral PCR products, wild-type vector PCR products that were left uncleaved by *Asc*I digestion were extracted from polyacrylamide gels and cloned. When 12 separate clones that lacked *Asc*I sites were sequenced, all were found to have acquired an identical single base change—a base substitution at the U3–R boundary that we call +1G—but all were otherwise identical to the parental sequence. An additional 10 clones that retained the *Asc*I site after PCR amplification were sequenced, and all possessed the wild-type sequence throughout the analyzed region.

The +1G substitution fortuitously introduced a *Kas*I recognition site, and hence we could examine the frequency of +1G by restriction analysis. This *Kas*I restriction site, which was detectable by restriction analysis in ~5% of the wild-type viral PCR products, was not detectable among PCR products generated by amplification of plasmid DNA, thus demonstrating that +1G is not a PCR artifact that results when LTR sequences are amplified and suggesting that the +1G substitution arose during viral replication (Figure 4). To confirm the rates of product generation suggested by restriction analysis, the frequency of various classes of reverse transcription products was also examined among clones of progeny DNAs of the –1/–14 vector. Although the RNA form of this vector lacks the *Asc*I restriction site, an *Asc*I site would be generated during reverse transcription if first strand transfer occurred precisely at the U3–R junction and the resulting primer-terminal mismatch were extended. Thirty-seven of 44 clones analyzed had *Asc*I sites, indicating that strand transfer had occurred precisely at the RNA's 5′ end. This frequency roughly agrees with the amount of *Asc*I digestion observed with the PCR product of viral DNAs. The remaining seven of the 44 clones all possessed the *Kas*1 site diagnostic of the previously observed +1G base substitution. Products of premature strand transfer would be expected to lack both the *Asc*I and the *Kas*I sites; however, no clones that lacked both sites were found in this screen. Taken together, these results indicated that the +1G mutation occurred in ~5–10% of the progeny DNAs of the –1/–14 mutant vector and at a similar level among wild-type vector progeny. This frequency is ~1000-fold higher than estimates of average base substitution rates at single sites during one round of M-MuLV replication (Monk *et al.*, 1992; Preston and Dougherty, 1996).

To determine whether any mutations other that +1G

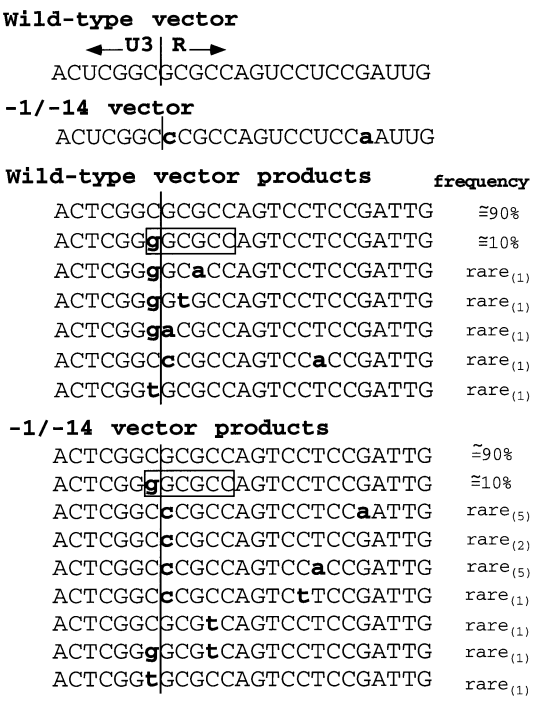

Fig. 5. Progeny DNA sequences. The sequences of the parental wild-type and –1/–14 vectors are presented at the top. For each of the two vectors, the first two product types listed were relatively common and their frequencies were approximated as described in the text. The remaining products were found rarely and were obtained by first depleting product pools of common products as described in the text. Numbers in parentheses given for each rare product indicate the number of individual subclones that were found to possess the indicated sequence. Differences between each product and the wild-type vector sequence are presented as bold lower case letters.

commonly occurred at the site of first strand transfer, we examined products that possessed neither the *Asc*I site diagnostic of precise first jump nor the *Kas*I site, which was indicative of the +1G mutation. Wild-type vector DNA products were depleted of these products by *Asc*I and *Kas*I digestion. Uncleaved products were gel purified, re-amplified, subcloned and sequenced. Roughly 10^4 reverse transcription products were used as starting material in this experiment: calculation of this number is based on the titer of puromycin-resistant colony formation and an assumption that half the viral DNA synthesized remains unintegrated (Barbosa *et al.*, 1994). Five clones that possessed neither an *Asc*I nor a *Kas*I site were sequenced and all were different from one another, suggesting that no single change other than +1G arises commonly during first strand transfer of the wild-type vector (Figure 5A). Analysis of –1/–14 vector products also failed to reveal any frequent changes at +1 other than +1G (Figure 5B). Note that because these DNAs had been subjected to as many as 60 cycles of PCR prior to sequencing, it is possible that some or all of the rarer sequence classes resulted from errors during *Taq* polymerase-directed synthesis rather than reverse transcriptase-directed errors during reverse transcription. As one means of examining the possible contribution of *Taq* polymerase errors to our study, sequencing was performed on an additional 14 clones which were derived from the same highly amplified pools that yielded the rare products in Figure 5 but that either retained the parental junction restriction site or else had a +1G-diagnostic *Kas*I site. In all of these, only one change in addition to those described below was detected: a single C→T substitution at –12 in one subclone that had retained the parental U3–R junction.

Although the +1G mutation appeared in ~10% of the products of a single round of reverse transcription, its relative abundance did not increase substantially in virus that had undergone many rounds of replication. This is evident from a comparison of the extent of *Asc*I cleavage of PCR products of serially passaged infectious virus DNA and of products of a single round of replication (Figure 4, lanes 7 and 8). Even alterations such as the deletion of one of the two U3 enhancer repeats, which would be expected to decrease transcription of a provirus harboring the deletion, accumulated during serial virus passage at a significantly higher rate than did +1G. This suggests that the +1G mutation must confer a disadvantage at some stage of replication and that most +1G observed among serially passaged virus products was newly formed rather than inherited from a parental provirus.

Frequency of premature strand transfer

Premature strand transfer was quite rare in our system. As described above, little evidence of premature strand transfer could be detected by restriction analysis of pooled viral DNA PCR products. Although this finding could have been due in part to partial restriction enzyme digestion, the failure to detect any clones resulting from premature strand transfer among 44 clones of –1/–14 vector products supported the notion that premature strand transfer was rare.

To examine further the rare premature strand transfer that did occur in our system, we analyzed pools of viral DNA products that had first been depleted of the commoner classes of reverse transcription products. Products of the –1/–14 mutant vector were depleted of the products containing a regenerated *Asc*I site or the +1G mutation by digestion with *Asc*I and *Kas*I. This reduced the pool to ~5% of its original size. The remaining uncleaved products were re-amplified and subcloned, and individual clones were analyzed for the presence of *Mfe*I (diagnostic of strand transfer prior to –14), *Kas*I or *Asc*I sites. Thirty-one of the 47 clones analyzed were found to have either an *Asc*I or a *Kas*I site and hence had resulted from incomplete digestion of the original PCR products. Another five clones had the *Mfe*I site diagnostic of strand transfer prior to –14. The remaining 11 were sequenced and the results are tabulated in Figure 5B. If all clones that contained a C at position –1 were assumed to have resulted from premature strand transfer, then the observed frequency of progeny DNAs suggests that strand transfer before position –14 and between positions –14 and –1 was about equally likely. Taken together, these data suggest that premature strand transfer (defined as strand transfer that occurred at least one nucleotide before the completion of –ssDNA) was involved in the synthesis of ~1–2% of the DNAs synthesized from the –1/–14 vector, an estimate roughly consistent with the frequency of *Mfe*I digestion visible among PCR products.

Testing possible causes of strand transfer-associated errors

Our finding that the first strand transfer occurred primarily from a single template position allowed us to use this

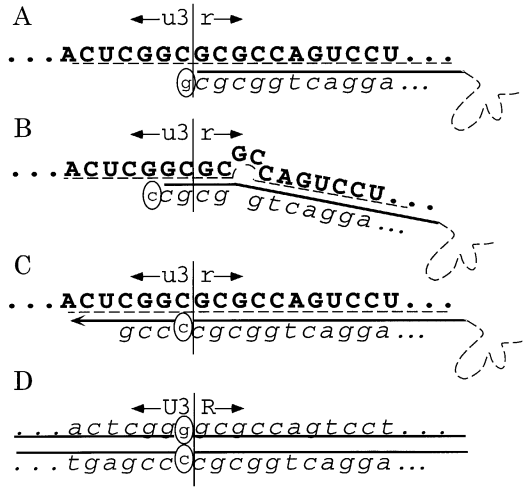

Fig. 6. Model for dislocation-mediated generation of the +1G mutation. (**A**) Correct alignment of –ssDNA with the wild-type vector acceptor template region and addition of the first post-jump nucleotide (circled). (**B**) Alternative misalignment of –ssDNA with the wild-type vector and template-directed addition of C (circled). (**C**) Re-alignment of the –ssDNA terminus with the acceptor template and extension of the terminal mispair. (**D**) Fixation of the dislocation-mediated substitution into product DNA upon completion of plus strand DNA synthesis.

system to study the errors which reverse transcriptase makes when it switches templates *in vivo*. To do this, we used specialized viral templates to test whether reverse transcriptase could use the same error mechanisms that it uses in cell-free systems during intracellular replication. We initially focused our experimentation on determining whether one of these mechanisms could account for the +1G substitution.

One common DNA polymerase error mechanism is simple base misincorporation. Reverse transcriptase occasionally incorporates template non-complementary bases, and hot spots for reverse transcriptase-mediated misincorporation have been observed (Bebenek and Kunkel, 1993). However, estimated rates of reverse transcriptase misincorporation are several orders of magnitude lower than the rate of +1G formation (Preston and Dougherty, 1996). Thus, we ruled that simple misincorporation was unlikely to cause +1G and did not test this possibility experimentally.

'Dislocation mutagenesis' is a class of DNA polymerase errors that arises through misalignment of the primer terminus (Kunkel, 1990) and that has been implicated in retroviral context-dependent hypermutation (Borman *et al.*, 1995). Dislocation mutagenesis involves template-directed incorporation from a misaligned primer terminus followed by re-alignment of the primer with the template and then extension of the resulting mismatch. The sequence of the wild-type vector at its u3–r junction is consistent with the possibility that a misaligned primer terminus could form upon strand transfer and cause reverse transcriptase to generate the +1G substitution (Figure 6).

To test whether +1G might result from dislocation, we generated M +1→+6, a vector that would be unable to generate +1G via the putative misaligned primer terminus proposed for the wild-type vector (Figure 2). If dislocation were the sole cause of +1G, this vector, which contained an *Mlu*I site in positions +6 through +1, would never template DNAs with G at position +1. To test this notion, reverse transcription products of the M +1→+6 vector that failed to be cut with *Mlu*I were subcloned and 11 were sequenced. Seven of these contained the +1G mutation in place of the terminal T of the parental vector's *Mlu*I site. This finding demonstrated that dislocation was not required to generate +1G.

Another possible origin of +1G was non-templated nucleotide addition followed by mispair extension. The only alteration we detected at high frequency at the strand transfer junctions of wild-type vector products was +1G. If the +1G mutation (a C→G substitution in the DNA plus strand) were caused by non-templated addition to –ssDNA before first strand transfer, then the non-templated nucleotide that was added to –ssDNA would have to have been a C. Such a finding would differ from observations in reconstituted reactions *in vitro*, where non-templated purines may be added more readily that pyrimidines (Patel and Preston, 1994). Alternatively, whereas the only change we detected among wild-type vector progeny was indicative of non-templated C addition, it was possible that a non-templated G might sometimes be added to –ssDNA. Non-templated G addition would not be detected because G would fortuitously be complementary to the acceptor template and hence product DNAs resulting from non-templated G addition prior to the first jump would not differ from the parental sequence.

To test the possibility that G was sometimes added as a non-templated nucleotide, we examined products of vectors with U at acceptor template position +1. Mismatch extension of the G–U base pair that would result if non-templated G were sometimes added to –ssDNA prior to the first jump would lead to +1C in the completed reverse transcription product. Consistent with the possibility of added G, all four clones of the 11 sequenced non-*Mlu*I site-containing M +1→+6 progeny DNAs that did not have +1G were found to possess +1C. Although this finding strongly supports the notion that non-templated G was added to –ssDNA prior to the first jump, the +1C substitution in the M +1→+6 progeny could conceivably have arisen due to an aberrant premature transcription start site, which could import an encoded +1C from the upstream LTR. Hence, although transcription initiation at pyrimidines is very rare, we tested the possibility that the +1C was caused by an alternate transcription start by constructing a proviral clone containing a T at position +1 in its upstream LTR as well as a T at position +1 in its downstream LTR. Reverse transcription products of vectors transcribed from this us+1; ds+1 (upstream +1 and downstream +1) mutant provirus would be predicted to contain a T at +1 regardless of the site of transcriptional initiation, provided reverse transcription were error-free. If, during reverse transcription, reverse transcriptase made a non-templated addition of G to –ssDNA, then the final outcome would be a T→C substitution at +1 in the plus strand of product DNA. Our us+1; ds+1 vector was designed so that a T→C substitution at +1 would generate an *Asc*I restriction site. When PCR-amplified reverse transcription products of the us+1; ds+1 vector were subjected to *Asc*I restriction analysis, 6% were found to possess a C at position +1 (Figure 7). Since this +1C would not have been present in either the donor or the acceptor template, this finding supports the notion that a

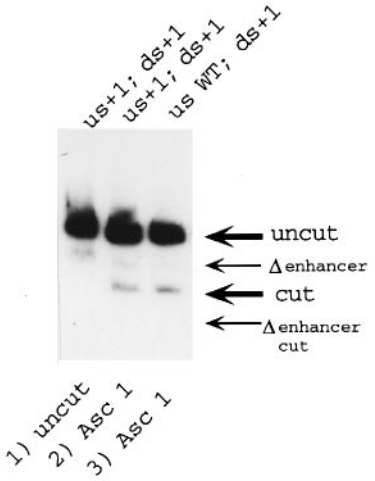

Fig. 7. Restriction analysis of +1U vector products. Radiolabeled PCR products of DNAs generated from vectors encoded by proviral DNAs containing +1T substitutions in both their upstream and downstream LTRs (vector is designated us+1; ds+1: lanes 1 and 2) or only their downstream LTRs (vector is designated us WT; ds+1: lane 3) are shown. Lane 1, undigested; lanes 2 and 3, digested with *Asc*I. Product mobilities and designations are as in Figure 3.

non-templated G was added to –ssDNA prior to the first strand transfer. A similar level of *Asc*I digestion was observed among products of a vector (us WT; ds+1) encoded by a plasmid containing the +1T substitution in only its acceptor r-encoding LTR, suggesting that little if any upstream initiation of transcription contributed to our findings.

Discussion

In this study, we examined the site of first strand transfer and the errors which reverse transcriptase makes during intracellular first strand transfer. We found that the first jump rarely takes place before the 5′ end of the RNA is reached, but that transfer at this site is highly error prone. All of the mutations we detected at the site of transfer were base substitutions: no insertions or deletions were observed.

Our results showed premature strand transfer one or more nucleotides before the completion of –ssDNA occurred during the synthesis of ~2% of progeny DNAs. However, even this low level may be an overestimate of premature strand transfer frequency. We calculated premature strand transfer rates using a vector with two single base substitutions. When amounts of products from virions harboring this vector were compared with those of the wild-type, we observed an ~10-fold decrease in DNA yield per unit virions (data not shown). We postulate that this decrease resulted from failure to extend –ssDNA–acceptor template mismatch. If so, then premature transfer products, which would have the same extent of template complementarity on both the wild-type and double mutant templates, would remain constant in amount on the two templates, and the entire observed decrease in product yield on the mutant template would come from products that were not premature. These considerations suggest that premature first strand transfer on the native template may be 10-fold lower than the 2% we observed, or as low as 0.2%.

Our findings regarding the site of first strand transfer conflict with reports in the literature but support prevailing models for reverse transcription. Our results are also consistent with unpublished results of J.Zhang and H.M.Temin, who used RNAs with r regions derived from two different viruses and who obtained results consistent with first strand transfer occurring at or near the 5′ end of RNA (J.Zhang, personal communication). Both previous studies that yielded results different from ours used linker insertion mutations to show that 5′ r mutations can be lost due to premature jumping (Lobel and Goff, 1985; Ramsey and Panganiban, 1993). We postulate that such template alterations may have adversely affected the vectors' replication competency and forced premature strand transfer. This suggestion is consistent with observations that some alterations to U5 decrease the ability of the vectors to serve as templates for –ssDNA (Jones *et al.*, 1994), and that some linker regions are hot spots for retroviral recombination (Pathak and Temin, 1990). One of the earlier strand transfer reports showed that RNAs with linker insertions generate short 'weak stop' DNAs in endogenous reactions in addition to –ssDNA (Lobel and Goff, 1985), an observation consistent with findings that template structure can interfere with reverse transcription (Klarmann *et al.*, 1993). Strand transfer has been proposed to occur via a pause and jump mechanism (Xu and Boeke, 1987; Telesnitsky and Goff, 1993) and a positive correlation has been observed between the frequency of pausing within a template region and the frequency of template switching within that region during reverse transcription in purified reactions (Wu *et al.*, 1995). The r linker insertions used in the previous reports may have provided an opportunity for reverse transcriptase to switch templates prematurely by creating a pause site before the end of the donor template was reached. In this work, we engineered mutations into only the acceptor r in order to minimize deleterious effects on transcription or the initiation of reverse transcription, and we used point mutations instead of linker insertion mutations.

We observed a high level of genetic variation at the first strand transfer site. Our assays analyzing the U3–R junction involved PCR amplification of viral sequences, and hence it is possible, especially for less frequent classes of products, that some of the mutations which we observed were caused by *Taq* polymerase or another enzyme that copied the viral sequences at some point. However, the high frequency of certain classes of alterations (e.g. +1G, which was found in ~10% of the PCR products of wild-type vector DNAs), paired with the absence of these sorts of alterations among serially amplified products of parental plasmid DNAs, supports the probability that the common classes of alterations we observed arose during retroviral replication. It is interesting to note that results with an HIV-based single replication cycle assay also show mutations at the U3–R junction, and evidence that strand transfer during yeast retrotransposition is error prone has been reported recently (Gabriel *et al.*, 1996; B.Preston, personal communication).

The results presented here support the model that +1 substitutions arise during reverse transcription via non-templated addition followed by mismatch extension upon strand transfer. Another possibility we considered was that +1G could potentially be templated by the 7-methyl-G

cap present on mRNAs and viral genomic RNAs (Coffin, 1996). Avian myeloblastosis virus reverse transcriptase can add a cap-complementary C residue during cDNA synthesis on mRNA *in vitro*, but not when the RNA has been de-capped (Volloch *et al.*, 1995). However, studies with purified enzymes and model primer–templates have demonstrated that +1G can arise at template switch junctions in the absence of a 7-methyl-G cap (Peliska and Benkovic, 1994), and our detection of +1C mutants demonstrates that not all additions to –ssDNA could be cap-templated.

The most frequent change we detected among wild-type vector products appeared to result from non-templated C addition and subsequent C–C mispair extension. In reconstituted reactions *in vitro*, the most commonly added non-templated nucleotides are purines, and C–C mispairs are extended particularly poorly by reverse transcriptases (Perrino, 1989; Ricchetti and Buc, 1990; Patel and Preston, 1994). However, studies *in vitro* have shown that the frequency with which nucleotides are added can differ from the frequency at which they are embedded into product DNAs. Furthermore, the rate at which mismatches are embedded does not appear to correlate with the ease of mispair extension, although some of the apparent paradox between what nucleotides become embedded and what nucleotides appear to be added preferentially may be a function of *in vitro* reaction conditions (Peliska and Benkovic, 1992, 1994; B.Preston, personal communication). In our studies, we were not measuring what nucleotides reverse transcriptase adds to –ssDNA but rather what nucleotides became embedded into product DNAs.

Reverse transcriptase template switches are not only necessary steps in the process of reverse transcription but are also critical in retroviral genetic recombination, since reverse transcriptase frequently performs template switches that can lead to retroviral recombination in addition to the two obligatory strand transfers (Hu *et al.*, 1993; Telesnitsky and Goff, 1993). It has been proposed that recombinogenic template switches may be highly mutagenic (Peliska and Benkovic, 1992; Patel and Preston, 1994; Wu *et al.*, 1995). This suggestion is based on the observation that mutations are very common in DNAs produced by purified reverse transcriptase that has been forced to switch templates (Peliska and Benkovic, 1994; Wu *et al.*, 1995). Like many other DNA polymerases, reverse transcriptase frequently will add an additional non-templated nucleotide when it reaches the end of a template *in vitro*: usually a purine (Clark, 1988). If this extended DNA switches to a secondary, acceptor template and synthesis continues, the nucleotide added at the end of the donor template can become fixed into the product DNA. Because the rate at which reverse transcriptase adds non-templated nucleotides in purified reactions is several logs higher than the rate of mismatch insertion at template-internal positions in reconstituted reactions *in vitro*, it has been suggested that recombination might generally be mutagenic (Patel and Preston, 1994). However, one study that involved sequencing 29 recombination junctions generated during intracellular reverse transcription revealed no template switch-associated mutations (Zhang and Temin, 1994), and another study that examined 18 recombination junctions found a mutation in only one (Stoye *et al.*, 1991). How do we reconcile our results with these previous findings? The authors of one study point out that template switching from template ends ('forced copy choice' recombination) might differ from switching from a template-internal region ('copy choice' recombination) (Zhang and Temin, 1994). However, studies in reconstituted model systems designed to examine copy choice-type recombination found low fidelity at template-internal switch junctions as well (Wu *et al.*, 1995). Hence, an alternate explanation for why our results are more similar to the observations from reconstituted reactions than to those of the previously studied virus-generated switch junctions may be that the *in vivo* switch junctions that were sequenced were a biased sample. When we forced mispair extension upon strand transfer with the –1 r mutant, the yield of product DNA dropped ~10-fold relative to that of the wild-type vector. These findings suggest it is possible that switch intermediates with added acceptor template non-complementary nucleotides tend to fail to complete recombinogenic template switches during viral replication, while those without acceptor template non-complementary nucleotides succeed. The unaltered junctions observed among *in vivo* products may thus reflect a bias among completed DNAs for those whose switch intermediates had had 3′ termini complementary to acceptor template regions. In contrast to recombinational template switches, which are not required in order to complete retroviral DNA synthesis, the first strand transfer is an obligatory step and all retroviral DNAs are the result of this strand transfer process. In the work presented here, we demonstrate that this first strand transfer during M-MuLV reverse transcription nearly always occurs from a single template position and that strand transfer at this position is highly error-prone. The question remains of whether template switching at any other single position, if forced to occur, would be as error prone as that reported here, or if this level of infidelity is a specific property of the first strand transfer.

Materials and methods

Plasmid construction

Tipless provirus plasmids. An *Eco*RI site was introduced into the middle of the U5 region of the downstream LTR of the replication-competent M-MuLV clone, pNCA (Colicelli and Goff, 1988), by standard PCR-mediated site-directed mutagenesis. Using this *Eco*RI site as one end and the *Nhe*I site in U3 that is 23 bp from the 'left' edge of M-MuLV's upstream LTR as the other end, 'tipless' virus-encoding sequences were subcloned into *Xba*I- plus *Eco*RI-cleaved pUC 18. Viral protein-encoding regions were then removed from this tipless provirus plasmid and replaced with the puromycin resistance gene by replacing sequences from the *Bsr*GI site that is early in *gag* to the *Cla*I site that is towards the end of *env* with the *Bsr*G1–*Cla*I puro resistance gene fragment of pBabe puro (Morgenstern and Land, 1990). The resulting plasmid contained the U3–R junction site of pNCA, which is different from the corresponding sequence in pBabe puro and the originally published sequence for M-MuLV (Shinnick *et al.*, 1981; Lobel and Goff, 1985). Tipless vector plasmids were used in the experiments presented here so that they would be distinguishable from reverse transcription products. These plasmids retained sufficient LTR sequences for expression of vector RNAs, but lacked intact LTRs. The progeny DNAs templated by these vectors had intact LTRs generated during reverse transcription.

Acceptor template region mutants. All mutations were introduced by standard PCR-mediated site-directed mutagenesis, confirmed by dideoxy sequencing, and introduced on *Cla*I–*Nhe*I restriction fragments into the puroR tipless provirus plasmid.

Cell and virus culture

Rat2 cells, NIH 3T3 cells, packaging cells and derivative producer cell lines were grown in Dulbecco's modified Eagle's medium supplemented with 10% calf serum (Gibco). The GP + E86 ecotropic packaging cell line (Markowitz *et al.*, 1988) was used with the permission of Genetix Pharmaceuticals, Inc. (Tarrytown, NY).

Stably transfected vector producer cells were established by transfecting GP + E86 cells with tipless provirus plasmids using the calcium phosphate method (Wigler *et al.*, 1979). Puromycin-resistant transfectants were selected by growth in 6 μg/ml puromycin (Sigma) and several hundred transfectants were pooled and used to produce virus for subsequent experimentation.

Viral supernatants were harvested from 90% confluent producer cells at 12 h intervals, filtered using 0.45 μM filters (Gelman) and stored at –80°C prior to use. In instances where virions were quantified, relative amounts of virions were determined by normalizing reverse transcriptase DNA polymerase activity levels in the culture media (Telesnitsky *et al.*, 1995). Infections were performed using thawed virus supernatants in the presence of 8 μg/ml hexadimethrine bromide (Sigma) for 2 h at 37°C.

High passage M-MuLV was infectious virus that had been passaged as cell-free virus for ~2 months. Initially the plasmid pNCA, which contains an infectious M-MuLV provirus (Colicelli and Goff, 1988), was transiently transfected by the DEAE–dextran method into NIH 3T3 cells (McCutchan and Pagano, 1968). Culture media containing viruses released over 12 h were harvested from 90% confluent cells, filtered and used undiluted to infect fresh 10% confluent 3T3 cells. Virus was harvested from these cells and serially passaged onto fresh cells as above.

Preparation and analysis of viral DNA

Low molecular weight cellular DNA enriched for viral sequences was extracted from Rat2 cells 15–20 h post-infection by the method of Hirt (1967). This DNA was amplified using U3 sense and U5 antisense PCR primers complementary to the right and left edges of an intact LTR, and *Taq* DNA polymerase from Perkin Elmer was used according to the manufacturer's directions. The primers were designed specifically to lie within sequences that were deleted from the tipless provirus expression plasmids but that were regenerated upon reverse transcription. PCR products were extracted with a 1:1 (v/v) mixture of phenol and chloroform, and the nucleic acids were collected by ethanol precipitation.

Where PCR products were subcloned, amplified sequences were digested with *Asp*718 (Boehringer Mannheim Biologicals) and *Xba*I and subcloned into pUC 18 without further purification. To deplete PCR products of species that contained a particular restriction enzyme site, a separate 5% polyacrylamide gel was used for each type of digested products. Digested products were subjected to electrophoresis, ethidium bromide stained, and bands visualized by UV light were excised with razor blades, placed in microcentrifuge tubes, covered with 100 μl of TE (10 mM Tris–Cl pH 8, 1 mM EDTA) and placed at 37°C for 24 h to allow passive product elution for use in subsequent re-amplification.

When products were to be quantified, one of the two PCR primers was radiolabeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. The unlabeled oligo was present at 1 μM and the labeled oligo was present at 0.2 μM; amplification was carried out for 25 cycles. Radiolabeled PCR products were diluted 2-fold with TE, extracted with equal volumes of phenol and chloroform, collected by ethanol precipitation in the presence of 100 mM NaCl, and digested with restriction enzymes (New England Biolabs) for 6–20 h using 20- to 40-fold excess enzyme and other reaction conditions as recommended by the manufacturer. Radiolabeled DNAs were subjected to electrophoresis on 5% polyacrylamide gels, and gels were dried and exposed to film. Quantification of products was performed using an Ambis Phosphorimaging system and/or quantification by liquid scintillation counting of excised gel slices.

DNA sequencing was performed using the Sequinase II kit (USB) and supercoiled plasmid miniprep templates. At least 70 nucleotides (from +35 to –35 relative to the U3–R junction) were sequenced unambiguously for each clone analyzed.

Acknowledgements

The authors thank Donna Ray for technical assistance, Caroline Kane and Millie Georgiadis for helpful discussion, Jiayou Zhang and Brad Preston for sharing unpublished results, and James Peliska, Mary Jane Wieland, David Friedman, Brad Preston and Caroline Kane for critical reading of the manuscript. This work was supported by American Cancer Society grant #VM-76657, with early stages supported in part by the John and Susan Munn Endowed Research Fund of the University of Michigan Comprehensive Cancer Center.

References

Barbosa,P., Charneau,P., Dumey,N. and Clavel,F. (1994) Kinetic analysis of HIV-1 early replicative steps in a coculture system. *AIDS Res. Hum. Retroviruses*, **10**, 53–59.

Bebenek,K. and Kunkel,T.A. (1993) The fidelity of retroviral reverse transcriptases. In Skalka,A.M and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 85–102.

Blain,S.W. and Goff,S.P. (1995) Effects on DNA synthesis and translocation caused by mutations in the RNase H domain of Moloney murine leukemia virus reverse transcriptase. *J. Virol.*, **69**, 4440–4452.

Borman,A.M., Quillent,C., Charneau,P., Kean,K.M. and Clavel,F. (1995) A highly defective HIV-1 group O provirus: evidence for the role of local sequence determinants in G→A hypermutation during negative-strand viral DNA synthesis. *Virology*, **208**, 601–609.

Clark,J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.

Coffin,J.M. (1979) Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.*, **42**, 1–26.

Coffin,J. (1996) Retroviridae: the viruses and their replication. In Fields,B.N, Knipe,D.M. and Howley,P.M. (eds), *Fundamental Virology.* 3rd edn. Lippincott-Raven, PA, pp. 763–843.

Coffin,J.M., Hageman,T.C., Maxam,A.M. and Haseltine,W.A. (1978) Structure of the genome of Moloney murine leukemia virus: a terminally redundant sequence. *Cell*, **13**, 761–773.

Colicelli,J. and Goff,S.P. (1988) Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.*, **199**, 47–59.

Das,A.T. and Berkhout,B. (1995) Efficient extension of a misaligned tRNA-primer during replication of the HIV-1 retrovirus. *Nucleic Acids Res.*, **23**, 1319–1326.

Gabriel,A., Willems,M., Mules,E.H. and Boeke,J.D. (1996) Replication infidelity during a single cycle of Ty1 retrotransposition. *Proc. Natl Acad. Sci. USA*, **93**, 7767–7771.

Gilboa,E., Mitra,S.W., Goff,S.P. and Baltimore,D. (1979) A detailed model of reverse transcription and tests of crucial aspects. *Cell*, **18**, 93–100.

Haseltine,W.A., Coffin,J.M. and Hageman,T.C. (1979) Structure of products of the Moloney murine leukemia virus endogenous DNA polymerase reaction. *J. Virol.*, **30**, 375–383.

Hirt,B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.*, **26**, 365–371.

Hu,W.-S., Pathak,V. and Temin,H.M. (1993) Role of reverse transcriptase in retroviral recombination. In Skalka,A.M and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 251–274.

Jones,J.S., Allan,R.W. and Temin,H.M. (1994) One retroviral RNA is sufficient for synthesis of viral DNA. *J. Virol.*, **68**, 207–216.

Klarmann,G.J., Schauber,C.A. and Preston,B.D. (1993) Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-I sequences *in vitro*. *J. Biol. Chem.*, **268**, 9793–9802.

Klaver,B. and Berkhout,B. (1994) Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res.*, **22**, 137–144.

Kunkel,T.A. (1990) Misalignment-mediated DNA synthesis errors. *Biochemistry*, **29**, 8003–8011.

Lobel,L.I. and Goff,S.P. (1985) Reverse transcription of retroviral genomes: mutations in the terminal repeats. *J. Virol.*, **53**, 447–455.

Markowitz,D., Goff,S. and Bank,A. (1988) A safe packaging line for gene transfer: separating viral genes on two different plasmids. *J. Virol.*, **62**, 1120–1124.

McCutchan,J.H. and Pagano,J.S. (1968) Enhancement of the infectivity of SV40 deoxyribonucleic acid with diethylaminoethyl-dextran. *J. Natl Cancer Inst.*, **41**, 351–357.

Monk,R.J., Mslik,F.G., Stokesberry,D. and Evans,L.H. (1992) Direct determination of the point mutation rate of a murine retrovirus. *J. Virol.*, **66**, 3683–3689.

Morgenstern,J.P. and Land,H. (1990) Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.*, **18**, 3587–3596.

Patel,P.H. and Preston,B.D. (1994) Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc. Natl Acad. Sci. USA*, **91**, 549–553.

Pathak,V.K. and Temin,H.M. (1990) Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc. Natl Acad. Sci. USA*, **87**, 6024–6028.

Peliska,J.A. and Benkovic,S.J. (1992) Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science*, **258**, 1112–1118.

Peliska,J.A. and Benkovic,S.J. (1994) Fidelity of *in vitro* DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Biochemistry*, **33**, 3890–3895.

Perrino,F.W., Preston,B.D., Sandell,L.L. and Loeb,L.A. (1989) Extension of mismatched 3′ termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl Acad. Sci. USA*, **86**, 8343–8347.

Preston,B.D. and Dougherty,J.P. (1996) Mechanisms of retroviral mutation. *Trends Microbiol.*, **4**, 16–21.

Pulsinelli,G.A. and Temin,H.M. (1994) High rate of mismatch extension during reverse transcription in a single round of retrovirus replication. *Proc. Natl Acad. Sci. USA*, **91**, 9490–9494.

Ramsey,C.A. and Panganiban,A.T. (1993) Replication of the retroviral terminal repeat sequence during *in vivo* reverse transcription. *J. Virol.*, **67**, 4114–4121.

Ricchetti,M. and Buc,H. (1990) Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent. *EMBO J.*, **9**, 1583–1593.

Roberts,J.D., Preston,B.D., Johnston,L.A., Soni,A., Loeb,L.A. and Kunkel,T. (1989) Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis *in vitro*. *Mol. Cell. Biol.*, **9**, 469–476.

Shinnick,T.M., Lerner,R.A. and Sutcliffe,J.G. (1981) Nucleotide sequence of Moloney murine leukemia virus. *Nature*, **293**, 543–548.

Stoye,J., Moroni,C. and Coffin,J.M. (1991) Virological events leading to spontaneous AKR thymomas. *J. Virol.*, **65**, 1273–1285.

Swanstrom,R., Varmus,H.E. and Bishop,J.M. (1981) The terminal redundancy of the retroviral genome facilitates chain elongation by reverse transcriptase. *J. Biol. Chem.*, **256**, 1115–1121.

Telesnitsky,A. and Goff,S.P. (1993) Strong-stop strand transfer during reverse transcription. In Skalka,A.M and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 49–83.

Telesnitsky,A., Blain,S. and Goff,S.P. (1995) Assays for retroviral reverse transcriptase. *Methods Enzymol.*, **262**, 347–362.

Temin,H.M. (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl Acad. Sci. USA*, **90**, 6900–6903.

Volloch,V.Z., Schweitzer,B. and Rits,S. (1995) Transcription of the 5′-terminal cap nucleotide by RNA-dependent DNA polymerase: possible involvement in retroviral reverse transcription. *DNA Cell Biol.*, **14**, 991–996.

Wigler,M., Sweet,R., Sim,G.K., Wold,B., Pellicer,A., Lacy,E., Maniatis,T., Silverstein,S. and Axel,R. (1979) Transformation of mammalian cells with genes from procaryotes and eucaryotes. *Cell*, **16**, 777–785.

Woehrl,B.M., Georgiadis,M.M., Telesnitsky,A., Hendrickson,W.A. and Le Grice,S.F.-J. (1995a) Footprint analysis of replicating murine leukemia virus reverse transcriptase. *Science*, **267**, 96–99.

Woehrl,B.M., Tantillo,C., Arnold,E. and Le Grice,S.F.J. (1995b) An expanded model of replicating human immunodeficiency virus reverse transcriptase. *Biochemistry*, **34**, 5343–5350.

Wu,W., Blumberg,B.M., Fay,P.J. and Bambara,R.A. (1995) Strand transfer mediated by human immunodeficiency virus reverse transcriptase *in vitro* is promoted by pausing and results in misincorporation. *J. Biol. Chem.*, **270**, 325–332.

Xu,H. and Boeke,J.D. (1987) High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **84**, 8553–8557.

Zhang,J. and Temin,H.M. (1994) Retrovirus recombination depends on the length of sequence identity and is not error prone. *J. Virol.*, **68**, 2409–2414.

Zinnen,S., Hsieh,J.C. and Modrich,P. (1994) Misincorporation and mispaired primer extension by human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.*, **269**, 24195–24202.

Received on July 30, 1996; revised on October 7, 1996